Unified Grid Tagging Scheme for Aspect Sentiment Quad Prediction

Guixin Su[1,2,3,4], Yongcheng Zhang[4], Tongguan Wang[1,2,3,4], Mingmin Wu[1,2,3,4], Ying Sha[1,2,3,4]*

[1]Key Laboratory of Smart Farming for Agricultural Animals, Wuhan, China

[2]Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education

[3]Hubei Engineering Technology Research Center of Agricultural Big Data, Wuhan, China

[4]College of Informatics, Huazhong Agricultural University, Wuhan, China

{cometsue,zhyc,wang_tg,wmm_nlp}@webmail.hzau.edu.cn, shaying@mail.hzau.edu.cn

Abstract

Aspect Sentiment Quad Prediction (ASQP) aims to extract all sentiment elements in quads for a given review to explain the reason for the sentiment. Previous table-filling based methods have achieved promising results by modeling word-pair relations. However, these methods decompose the ASQP task into several subtasks without considering the association between sentiment elements. Most importantly, they fail to tackle the situation where a sentence contains multiple implicit expressions. To address these limitations, we propose a simple yet effective **U**nified **G**rid **T**agging **S**cheme (**UGTS**) to extract sentiment quadruplets in one shot, with two additional special tokens from pre-trained models to represent potential implicit aspect and opinion terms. Based on this, we first introduce the adaptive graph diffusion convolution network to construct the direct connection between explicit and implicit sentiment elements from syntactic and semantic views. Next, we utilize conditional layer normalization to refine the mutual indication effect between words for matching valid aspect-opinion pairs. Finally, we employ the triaffine mechanism to integrate heterogeneous word-pair relations to capture higher-order interactions between sentiment elements. Experimental results on four benchmark datasets show the effectiveness and robustness of our model, which achieves state-of-the-art performance.

1 Introduction

As a fine-grained sentiment analysis task (Pontiki et al., 2014), Aspect Sentiment Quad Prediction (ASQP) aims to identify sentiment quadruplets from a given review, to explain what the targeted aspects are, which domain described (aspect category), how their sentiment polarities are and why they have such polarities (opinion terms). It is challenging as aspects and opinions can be explicitly expressed or implicitly conveyed. Take the example in Figure 1 as the task illustration, the sentence contains four different types of sentiment quadruplets with explicit and implicit aspects/opinions.

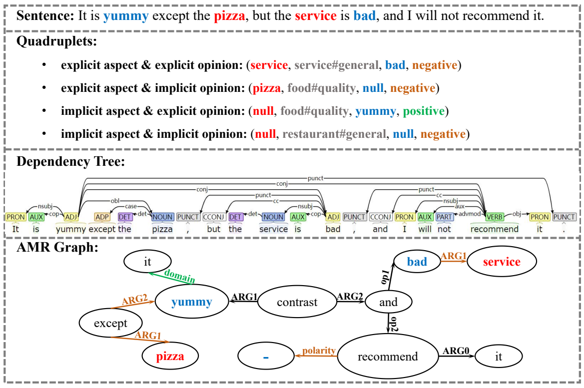

Figure 1: A sentence with its dependency tree and abstract meaning representation (AMR) graph is given to illustrate the ASQP task. It reveals the specific relations between matched aspects and opinions in the syntactic and semantic view. In the quadruplet set, aspect term, aspect category and opinion term are marked in red, gray and blue, respectively. The positive sentiment polarity is highlighted in green, while the negative is in brown. Note that implicit aspect and opinion are set as "null".

Indeed, ASQP can be decomposed into several subtasks, such as Aspect Term Extraction (ATE) (Li et al., 2020; Wang et al., 2021), Opinion Term Extraction (OTE) (Fan et al., 2019; Wu et al., 2020b), Aspect-based Sentiment Classification (ASC) (Tang et al., 2022; Zhang et al., 2023), Aspect Category Detection (ACD) (Hu et al., 2021; Nguyen et al., 2023), Aspect-Opinion Pair Extraction (AOPE) (Chen et al., 2020; Gao et al., 2021), and Aspect Sentiment Triplet Extraction (ASTE) (Peng et al., 2020; Su et al., 2024), etc. Therefore, early work adopts a two-stage pipeline approach to first extract the candidate aspect-opinion pairs, and then predict the aspect category and sentiment polarity (Cai et al., 2021). However, pipeline methods easily suffer from the error propagation problem.

*Corresponding author.

Proceedings of the 31st International Conference on Computational Linguistics, pages 3997–4010
January 19–24, 2025. ©2025 Association for Computational Linguistics

To alleviate this problem, many efforts resort to extracting sentiment quadruplets in an end-to-end framework. Some works formulate the ASQP task as a generative problem, leveraging generative pre-trained models like BART (Lewis et al., 2020) and T5 (Raffel et al., 2020) to generate sentiment quadruplets in one shot (Zhang et al., 2021a; Mao et al., 2022; Gou et al., 2023). In addition, some methods employ the data augmentation strategy to address the data scarcity problem (Wang et al., 2023; Yu et al., 2023; Zhang et al., 2024a,b). However, these approaches neglect to establish reciprocity between sentiment elements within the quadruplets, which is the key to resolving the ASQP task.

Therefore, some works model such association by tagging the word-pair relations. Zhou et al. (2023) propose a sentiment-specific horns tagging schema to jointly identify aspect category and aspect-opinion-sentiment triplet. However, this method ignores the higher-order interactions between aspect category and other sentiment elements. And the "null" token added at the beginning of a sentence may distort the semantic expression in some cases. Most importantly, it cannot solve the case of the implicit aspect with implicit opinion. Meanwhile, Zhu et al. (2023) leverage a sentence-guided grid tagging scheme to capture both explicit and implicit expressions in sentences. But it struggles to simultaneously tackle the situation where a sentence with multiple implicit expressions.

To address the above issues, we introduce a simple yet effective **U**nified **G**rid **T**agging **S**cheme (**UGTS**) to transform the ASQP task into a unified table-filling task by constructing the word-pair relations. UGTS incorporates two special tokens from pre-trained models (e.g., [CLS] and [SEP] for BERT) at the beginning and end of the sentence to respectively represent the potential implicit aspect and opinion term, without distorting the semantics of the sentence. It can not only extract sentiment quadruplets in a unified fashion without decomposing the ASQP task into several subtasks, but also tackle concurrent situations where multiple or multi-type quadruplets exist.

Based on this, we leverage syntactic and semantic structure information derived from the dependency tree and Abstract Meaning Representation (AMR) to enhance the association between sentiment elements. The dependency tree illustrates the syntactic dependencies between words, while the AMR captures the semantic dynamics of "who is doing what to whom" among arguments. These linguistic features help reveal the specific relations between the matched aspects and opinions. Take Figure 1 as an example, aspect "service" is the nominal subject of the opinion "bad" in the syntactic view, and aspect "service" serves as the proto-patient of the opinion "bad" in the semantic view.

To fully exploit these linguistic features, we first introduce an adaptive graph diffusion convolution network to establish the connection between explicit and implicit sentiment elements, as the dependency tree and AMR graph only parse the original sentence, ignoring the two special tokens denoting implicit aspect and opinion. Secondly, we utilize conditional layer normalization to refine the mutual indication effect between aspect-opinion pairs, as aspects provide aspect category information for opinions and conversely convey sentiment polarity information. Finally, we employ the triaffine mechanism to align the syntactic and semantic aware word-pair relations into the context, and then aggregate them to capture higher-order interactions between sentiment elements.

In summary, the key contributions are as follows:

- To the best of our knowledge, we make the first effort to introduce a Unified Grid Tagging Scheme (UGTS) to address ASQP in a unified table-filling task, without decomposing the ASQP task into several subtasks. Besides, UGTS can tackle concurrent situations where multiple or multi-type quadruplets exist.
- Based on the UGTS, we propose a novel end-to-end model to fully develop the syntactic and semantic features derived from the dependency tree and AMR graph, to further enhance the association between sentiment elements.
- We conduct extensive experiments on four benchmark datasets, and the experimental results show the effectiveness of our model.

2 Related Works

Aspect Sentiment Quad Prediction (ASQP) (Zhang et al., 2021a), also known as Aspect Category Opinion Sentiment Quadruple Extraction (ACOS) (Cai et al., 2021), has attracted substantial research in recent years, which can be categorized into the follow three main kinds.

Sequence tagging. Studies have focused on employing the Begin-Inside-Outside tagging scheme to perform sequence labeling to extract aspects and opinions, then predict the sentiment polarity and aspect category of the valid aspect-opinion pairs. Cai et al. (2021) propose a two-stage pipeline approach to perform Cartesian Product to obtain candidate aspect-opinion pairs and then assign the category-sentiment class. Similarly, Zhang et al. (2021a) benchmark ASQP task with several baselines, which decompose the ASQP task into multiple subtasks to extract the sentiment elements successively and then group them into quadruplets. Obviously, these methods potentially lead to the well-known error propagation problem.

Generative based. Generative methods transform the ASQP task into a text generation problem, which utilize the unified frameworks to directly generate the label sequence or desired sentiment elements given the input sentence (Peper and Wang, 2022; Hu et al., 2023; Li et al., 2023; Kim et al., 2024). These methods can mainly be classified as template-based (Zhang et al., 2021a; Hu et al., 2022; Gao et al., 2022; Gou et al., 2023; Mohammadkhani et al., 2024) and structure-based (Mao et al., 2022; Bao et al., 2022, 2023b,a). Despite achieving promising results, these methods exhibit differences between the training and inference stages, causing the exposure bias problem.

Data augmentation. Studies focus on expanding the number of training samples to solve the data scarcity problem owing to the high annotation cost. Wang et al. (2023) train a quads-to-text model to enrich the semantic diversity of the generated data and design a data filtering strategy to remove low-quality augmented data. To reduce mismatches in data augmentation, Yu et al. (2023) and Zhang et al. (2024b) utilize the self-training mechanism to obtain high-quality samples. In addition, Zhang et al. (2024a) present an adaptive data augmentation framework to tackle the quad-pattern imbalance and aspect-category imbalance issue. However, these approaches are constrained in capturing reciprocity between sentiment elements. Therefore, we propose a unified grid tagging scheme to model such association.

3 Unified Grid Tagging Scheme

In this section, we first introduce the task definition of Aspect Sentiment Quad Prediction (ASQP) and then illustrate how to represent the ASQP task in our proposed unified grid tagging scheme. Finally, we present the quadruplet decoding algorithm to parse the tagging results.

Tag	Meaning
B-A	Beginning of the aspect term
I-A	Inside of the aspect term
B-O	Beginning of the opinion term
I-O	Inside of the opinion term
A	Two words belong to the same aspect term
O	Two words belong to the same opinion term
POS NEU NEG c_i	Two words respectively belong to an aspect and an opinion, and they form a sentiment quadruplet with positive/neutral/negative sentiment polarity and c_i aspect category, where $c_i \in \mathcal{C} = \{c_1, c_2, \cdots, c_m\}$.
N	No relation between two words as described above

Table 1: The meanings of word-pair relations for the ASQP task.

3.1 Problem Formulation

Given an input sentence $X = \{w_1, w_2, \cdots, w_n\}$ consisting of n words, the objective of the ASQP task is to extract all possible aspect-level sentiment quadruplets $\mathcal{Q} = \{(at, ac, ot, sp)_k\}_{k=1}^{|\mathcal{Q}|}$ corresponding to four sentiment elements. In particular, the aspect term (at) and the opinion term (ot) are normally continuous spans in the sentence to explicitly express the opinion target and the subjective statement. Note that they are set as "null" in case of implicit mention. The aspect category (ac) describes a specific domain of interest for the aspect term within the predefined label set $\mathcal{C} = \{c_1, c_2, \cdots, c_m\}$ with m categories. The sentiment polarity (sp) belongs to the sentiment class set $\mathcal{S} = \{POS, NEU, NEG\}$ denoting the positive, neutral and negative semantic orientation, respectively. The sentence X has a total number of $|\mathcal{Q}|$ sentiment quadruplets.

3.2 Tagging Scheme Description

Inspired by Wu et al. (2020a) and Chen et al. (2022), we propose a simple yet effective unified grid tagging scheme to model the association between sentiment elements, to formulate the ASQP task as a table-filling task, without decomposing the ASQP task into several subtasks to extract different sentiment elements separately.

The basic idea of the unified grid tagging scheme is to tag a word-pair table T with $|n+2| \times |n+2|$ grids, where n is the length of the input sentence and 2 denotes the special tokens from the pre-trained language model (e.g., [CLS] and [SEP] for BERT). Note that we add these two special tokens at the beginning and end of the sentence

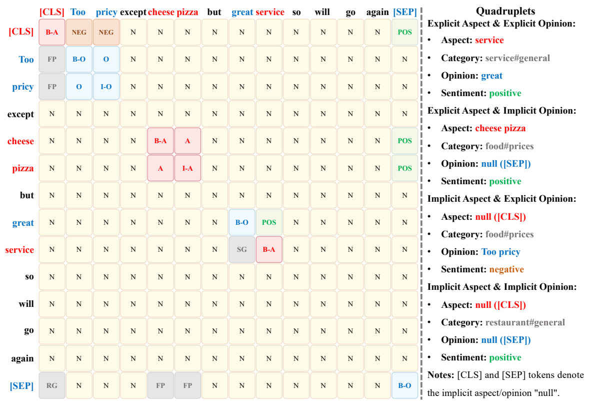

Figure 2: A tagging example with different types of sentiment quadruplets for ASQP task.

to represent the potential implicit aspect term and opinion term, respectively. For each grid, $T[i][j]$ corresponds to the word-pair relation between the i-th word and the j-th word.

To comprehensively consider all possible associations between sentiment elements, we define $10 + m$ types of word-pair relations for ASQP, where m denotes the number of predefined aspect categories varying for different datasets. The meanings of tags are elaborated in Table 1, and a corresponding unified grid tagging example is illustrated in Figure 2.

Specifically, in the main diagonal region, we utilize boundary-driven tags {B-A, I-A, B-O, I-O} to determine the beginning and inside of aspect and opinion terms. Note that if the special tokens [CLS] and [SEP] are marked as "B-A" and "B-O" tags respectively, it can be considered that the sentence contains implicit aspect term and opinion term. In the non-diagonal region, we employ tags {A, O} to detect whether the word pair formed by two different words belongs to the same aspect or opinion term, respectively. Meanwhile, we adopt the sentiment tags {POS, NEU, NEG} in the upper triangular region and the aspect category tags $c_i \in \mathcal{C}$ in the lower triangular region, not only to match the valid aspect-opinion pairs but also to judge their sentiment polarities and aspect categories. The "N" tag indicates that there is no relation between words as described above.

In short, the unified grid tagging scheme can: (1) effectively construct the association between sentiment elements. (2) add two special tokens denoting the implicit aspect and opinion term without affecting the semantics of the sentence. (3) simultaneously tackle the concurrent situations of multiple or multi-type quadruplets. (4) directly address several subtasks without additional modifications.

Algorithm 1 Quadruplet decoding for ASQP task

Input: The prediction results $P = \{p_{00}, p_{01}, \cdots, p_{n+1,n+1}\}$ of the sentence X (n words) with two additional special tokens. p_{ij} denotes the predicted tag of the word pair (w_i, w_j).
Output: Quadruplets set $\mathcal{Q}$ of the given sentence.

Initialize the $A = \{\}, O = \{\}, \mathcal{Q} = \{\}$
if p_{00} = "B-A" **then** $A.append$([CLS])
if $p_{n+1,n+1}$ = "B-O" **then** $O.append$([SEP])
while left index $1 \le l \le n$ and right index $l \le r \le n$ **do**
  if p_{ll} = "B-A" meanwhile p_{kk} = "I-A" when $l < k \le r$ **then** $A.append([w_l, w_{l+1}, \cdots, w_r])$
  if p_{ll} = "B-O" meanwhile p_{kk} = "I-O" when $l < k \le r$ **then** $O.append([w_l, w_{l+1}, \cdots, w_r])$
end while
while $a \in A$ and $o \in O$ **do**
  $S = \{\}, C = \{\}$
  while $w_i \in a$ and $w_j \in o$ **do**
    if $i < j$ **then** $stag = p_{ij}$, $ctag = p_{ji}$
    else $stag = p_{ji}$, $ctag = p_{ij}$
    if $stag \in \{POS, NEU, NEG\}$ and $ctag \in \mathcal{C}$ **then**
      $S.append(stag), C.append(ctag)$
  end while
  if $S \ne \emptyset$ and $C \ne \emptyset$ **then** $s = argmax(S)$, $c = argmax(C)$, $\mathcal{Q}.append(a, c, o, s)$
end while

3.3 Quadruplet Decoding

The details of sentiment quadruplet decoding are shown in Algorithm 1. Briefly, from the main diagonal of the prediction results, we first recognize aspect and opinion terms based on a span consisting of consecutive "B-A" and "I-A" tags or "B-O" and "I-O" tags. Note that implicit aspect and opinion term are detected when the special tokens [CLS] and [SEP] are labeled as "B-A" and "B-O" tag respectively. Next, we need to determine whether the aspect terms and opinion terms match. If there exists any sentiment relation and aspect category between them, they are considered as the matched aspect-opinion pair. Finally, for each matched aspect-opinion pair, we identify the sentiment polarity by selecting the most predicted sentiment relation in the upper triangular region. A similar operation yields the aspect category from the prediction results in the lower triangular region. Therefore, we collect each sentiment quadruplet and construct the predicted quadruplet set.

4 Methodology

In this section, we first elaborate on the details of our proposed model and then describe the training objective. The overall architecture of our model is illustrated in Figure 3.

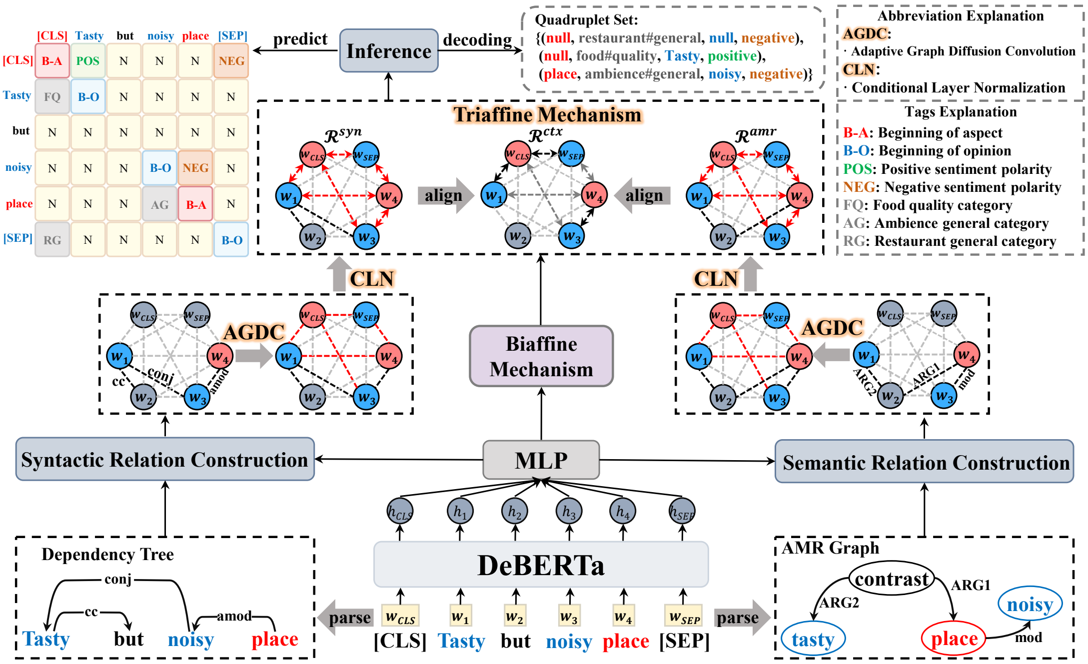

Figure 3: The overall architecture of our model.

4.1 Input and Encoding Layer

BERT-style models (Devlin et al., 2019) have demonstrated their effectiveness in various natural language processing tasks. In this work, we utilize DeBERTa (He et al., 2021) as the context encoder to encode the given sentence $X = \{w_1, w_2, \cdots, w_n\}$ added with [CLS] and [SEP] tokens. The encoding layer outputs the hidden contextual features $\mathcal{H} = \{h_{cls}, h_1, \cdots, h_n, h_{sep}\} \in \mathbb{R}^{(n+2)\times d_h}$ at the last Transformer block.

4.2 Relation Construction

To effectively construct the association between sentiment elements, our model adopts the aforementioned unified grid tagging scheme as the modeling paradigm. The core lies in deeply exploring the relational features between words, which can reveal the interactions between sentiment elements to facilitate the sentiment quadruplets extraction.

Therefore, we leverage syntactic and semantic structure information derived from the dependency tree and Abstract Meaning Representation (AMR) to enhance the representation of word-pair relations, as these linguistic features effectively reveal specific relations between aspect-opinion pairs. For example, aspects typically manifest as the nominal subject of opinions in the dependency tree and as the proto-patient in the AMR graph. These relational features provide the model with insight into the interactions between sentiment elements.

Specifically, we utilize the off-the-shelf toolkits spaCy[1] and AMRLib[2] to parse sentences to obtain corresponding syntactic dependency trees and AMR graphs. Note that we further align the AMR graphs with the original sentences by the aligner Penman[3] (Goodman, 2020) to ensure that the arguments in the AMR graph are logically mapped to the words in the sentence. Based on this, we can obtain various relations of word pairs from syntactic and semantic views. For each linguistic feature, it is intuitive to construct as a word-pair relation graph, where each word and word-pair relation is regarded as node and edge, respectively. In particular, we add a "self" relation type to represent the self-loop edge for each word itself, and assign the "none" type if there is no relation between two words. For feature initialization, we employ the hidden contextual representations $\mathcal{H}$ as node features, and represent edge features as an adjacency matrix $\mathcal{R} = \{r_{ij} \mid 0 \leq i, j \leq n+1\}$, where $r_{ij} \in \mathbb{R}^{d_p}$ is the embedding of the edge label for word pair (w_i,w_j) by looking up a trainable embedding table, and d_p is number of word-pair tags.

4.3 Adaptive Graph Diffusion Convolution

Next, we introduce the adaptive graph diffusion convolution network (AGDC) to establish the di-

[1] https://spacy.io/
[2] https://amrlib.readthedocs.io/en/latest/
[3] https://github.com/goodmami/penman

rect connection between explicit and implicit sentiment elements in the syntactic and semantic aware word-pair relation graphs, as [CLS] and [SEP] tokens (denoting implicit aspect and opinion) are not involved in the syntax and AMR parsing.

Specifically, for syntactic and semantic aware word-pair relation graphs, we first utilize the adaptive graph diffusion strategy based on heat kernel weight coefficients to support learning the optimal neighborhood of different word nodes. The strategy can capture the multi-hop associations between word nodes to construct the direct connection between explicit and implicit sentiment elements. The process is represented as:

$$\hat{\mathcal{R}}^{syn}; \hat{\mathcal{R}}^{amr} = \sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} (\mathcal{R}^{syn/amr})^k \mathcal{H} \quad (1)$$

where $t \in \mathbb{R}^{n+2}$ denotes the trainable neighborhood radius for different words, $\mathcal{R}^{syn}$ and $\mathcal{R}^{amr}$ are the initialized edge features of syntactic and semantic aware word-pair relation graphs, and $\mathcal{H}$ is the initialized node features. Then we employ a graph convolution network (Kipf and Welling, 2017) to aggregate the associated opinion information for aspect terms to facilitate their matching:

$$\hat{\mathcal{H}}^{syn}; \hat{\mathcal{H}}^{amr} = \mathrm{ReLU}(\hat{\mathcal{R}}^{syn/amr} \mathcal{H} W + b) \quad (2)$$

where W and b are the learnable weight and bias.

4.4 Conditional Layer Normalization

Intuitively, aspects provide aspect category information for opinions, while opinions convey sentiment polarity. Thus, it is vital to refine the mutual indication effect between aspects and opinions to facilitate the sentiment quadruplets extraction.

Inspired by Liu et al. (2021), we adopt the conditional layer normalization (CLN) to model the directed word-pair relation between words:

$$\tilde{r}_{ij} = \mathrm{CLN}(\hat{h}_i, \hat{h}_j) = \gamma_{ij} \odot (\frac{\hat{h}_j - \mu}{\sigma}) + \lambda_{ij} \quad (3)$$

where $\hat{h}_i$ is the condition to determine the scaling factor $\gamma_{ij} = W_\gamma \hat{h}_i + b_\gamma$ and bias $\lambda_{ij} = W_\lambda \hat{h}_i + b_\lambda$ of layer normalization. μ and σ are the mean and standard deviation across the elements of $\hat{h}_j$, i.e.,

$$\mu = \frac{1}{d_h} \sum_{k=1}^{d_h} \hat{h}_{jk}, \quad \sigma = \sqrt{\frac{1}{d_h} \sum_{k=1}^{d_h} (\hat{h}_{jk} - \mu)^2} \quad (4)$$

where $\hat{h}_{jk}$ is the k-th dimension of $\hat{h}_j$. We apply CLN to transform node features $\hat{\mathcal{H}}^{syn}$ and $\hat{\mathcal{H}}^{amr}$ to obtain refined word-pair relations $\tilde{\mathcal{R}}^{syn}$ and $\tilde{\mathcal{R}}^{amr}$ from syntactic and semantic views.

4.5 Triaffine Mechanism

The refined syntactic and semantic aware word-pair relations reveal the mutual indication effect between aspects and opinions from their unique views. To ensure that these heterogeneous features complement and reinforce each other within a unified feature space, it is necessary to further align and aggregate them to capture the higher-order interactions between sentiment elements.

Specifically, since biaffine mechanism (Dozat and Manning, 2017) has been proven effective in modeling the word-pair interactions, we utilize it to convert the hidden states $\mathcal{H}$ into contextual feature space $\mathcal{R}^{ctx} = \{r_{ij}^{ctx} \mid 0 \leq i, j \leq n+1\}$, i.e.,

$$r_{ij}^{ctx} = \begin{bmatrix} h_j \\ 1 \end{bmatrix}^{\mathrm{T}} W_{ba} h_i \quad (5)$$

where $r_{ij}^{ctx} \in \mathbb{R}^{d_p}$ is the contextual relation for word pair (w_i, w_j), and W_{ba} is the trainable weight.

Next, inspired by (Yuan et al., 2022) in integrating multi-source heterogeneous features, we employ the triaffine mechanism to project the refined syntactic and semantic aware word-pair relations $\tilde{\mathcal{R}}^{syn}$ and $\tilde{\mathcal{R}}^{amr}$ into the contextual feature space $\mathcal{R}^{ctx}$, thereby aligning specific relations between sentiment elements revealed in the dependency tree and AMR graph. The operation is formulated as:

$$p_{ij} = \begin{bmatrix} \tilde{r}_{ij}^{syn} \\ 1 \end{bmatrix}^{\mathrm{T}} (r_{ij}^{ctx})^{\mathrm{T}} W^{tri} \begin{bmatrix} \tilde{r}_{ij}^{amr} \\ 1 \end{bmatrix} \quad (6)$$

where $p_{ij} \in \mathbb{R}^{d_p}$ is the syntactic and semantic enhanced representations of word pair (w_i, w_j) for prediction as logits, and W_{tri} is the trainable weight.

4.6 Training Objective

The tagging grids actually contain many "N" labels irrelevant to sentiment elements, which leads to the class imbalance problem of word-pair relations and misguides the parameter optimization. Therefore, we employ Focal loss (Lin et al., 2017) as training loss to mitigate the imbalance problem, i.e.,

$$\log(p_t) = \sum_{i=0}^{n+1} \sum_{j=0}^{n+1} \sum_{z \in \mathcal{Z}} \mathbb{I}(y_{ij} = z) \log(p_{ij|z}) \quad (7)$$

$$\mathcal{L} = -\alpha_t (1 - p_t)^\beta \log(p_t) \quad (8)$$

where $\mathbb{I}(\cdot)$ is the indicator function, y_{ij} is the ground truth of word pair (w_i, w_j), and $\mathcal{Z}$ signifies pre-defined tags. Hyperparameter α_t is a factor to balance between the positive and negative samples, and β is the focusing parameter that regulates the weights of easily categorized samples.

Methods	Laptop			Rest			Rest15			Rest16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Extract-Classify	45.56	29.28	35.80	38.54	52.96	44.61	35.64	37.25	36.42	38.40	50.93	43.77
ChatGPT	21.72	27.65	24.33	38.39	46.40	42.02	29.66	37.86	33.26	36.09	46.93	40.81
Paraphrase	-	-	-	-	-	-	46.16	47.72	46.93	56.63	59.30	57.93
SGTS-ASQE	41.44	32.26	36.28	55.24	43.69	48.76	-	-	-	-	-	-
Seq2Path	-	-	42.97	-	-	58.41	-	-	-	-	-	-
GAS	43.46	42.69	43.07	59.81	57.51	58.63	47.15	46.01	46.57	57.30	57.82	57.55
ILO	44.14	44.56	44.35	58.43	58.95	58.69	47.78	50.38	49.05	57.58	61.17	59.32
DLO	43.40	43.80	43.60	60.02	59.84	59.18	47.08	49.33	48.18	57.92	61.80	59.79
SS-UAUL	44.38	43.65	44.01	61.22	59.87	60.53	49.12	50.39	49.75	59.24	61.75	60.47
One-ASQP	43.80	39.54	41.56	65.91	56.24	60.69	-	-	-	-	-	-
MvP	-	-	43.92	-	-	61.54	-	-	51.04	-	-	60.39
ADA	45.03	44.53	44.78	60.15	61.95	61.04	49.31	53.96	51.53	59.34	62.83	61.03
MUL-ST-Scorer	47.05	45.32	<u>46.17</u>	65.43	61.92	<u>63.63</u>	51.94	52.00	<u>51.97</u>	63.46	64.31	<u>63.88</u>
Ours	48.21	46.39	**47.28**	65.94	63.47	**64.68**	52.76	52.43	**52.59**	65.72	64.50	**65.10**

Table 2: Experimental results on four datasets (%). The best results are in **bold** and the second best are <u>underlined</u>.

5 Experiments

5.1 Datasets and Setup

We evaluate our model on four benchmark ASQP datasets released by Cai et al. (2021)[4] and Zhang et al. (2021a)[5]. These datasets are derived from SemEval Challenges (Pontiki et al., 2015, 2016) and Amazon platform during 2017 and 2018, with one in the laptop domain and three in the restaurant domain. Detailed statistics about datasets are shown in **Appendix** A.1. Following previous works, we employ F1 scores (F1) as the main evaluation metric and also report the corresponding Precision (P) and Recall (R) scores. Implementation details are presented in **Appendix** A.2.

5.2 Baselines

We compare our model with the following three main kinds of state-of-the-art baselines: **1) Tagging based methods:** Extract-Classify (Cai et al., 2021), SGTS-ASQE (Zhu et al., 2023), One-ASQP (Zhou et al., 2023). **2) Generative methods:** ChatGPT (Xu et al., 2023), Paraphrase (Zhang et al., 2021a), Seq2path (Mao et al., 2022), GAS (Zhang et al., 2021b), ILO/DLO (Hu et al., 2022), SS-UAUL (Hu et al., 2023), MvP (Gou et al., 2023). **3) Data augmentation:** ADA (Zhang et al., 2024a), MUL-ST-Scorer (Zhang et al., 2024b). The details about baselines are elaborated in the Related Works.

5.3 Main Results

The main results are reported in Table 2. Overall, our model achieves superior performance over all state-of-the-art baselines under the F1 scores.

Specifically, we have the following observations: (1) End-to-end methods achieve more significant improvements than the pipeline method Extract-Classify, as the former establishes the correlations between subtasks by jointly training them to alleviate the error propagation problem. (2) Compared with generative methods, our model significantly improves F1 scores by an average of 3.63% and 5.18% on Laptop and Rest, and achieves impressive increase by an average of 4.0% and 5.89% on Rest15 and Rest16. Because generative methods overlook the reciprocity among the sentiment elements. (3) Our model exceeds data augmentation methods by an average of 1.0%~2.82% F1 scores. This improvement is attributed to that our model considers the higher-order interactions between sentiment elements. (4) Note that our model outperforms table-filling based methods SGTS-ASQE and One-ASQP by a large margin. We suppose the reason is that our unified grid tagging scheme can simultaneously tackle the situation where a sentence with multiple implicit expressions. Besides, our model leverages syntactic and semantic linguistic features to enhance word-pair relations.

5.4 Ablation study

To verify the effectiveness of different modules in our model and the rationality of using different linguistic features, we conduct an ablation study and the experimental results are shown in Table 3.

Specifically, for module ablation, **w/o AGDC** denotes we remove the adaptive graph diffusion convolution module and achieves the dropping performance with decreasing 3.53% F1 scores on aver-

[4] https://github.com/NUSTM/ACOS
[5] https://github.com/IsakZhang/ABSA-QUAD

Model	Laptop	Rest	Rest15	Rest16
Full model	**47.28**	**64.68**	**52.59**	**65.10**
w/o AGDC	43.55	60.71	49.37	61.92
w/o CLN	44.69	61.97	50.43	62.47
w/o Triaffine	45.81	63.02	51.24	63.42
w/o AMR	42.60	59.83	47.68	60.19
w/o Syn	41.53	60.14	46.91	60.21
w/o Syn+AMR	39.68	56.52	43.82	56.43

Table 3: Ablation study (average F1 reported).

Dataset	Model	Quadruplet types			
		EA&EO	IA&EO	EA&IO	IA&IO
Laptop	Extract-Classify	35.39	39.00	16.82	18.58
	SGTS-ASQE	36.58	52.45	17.24	14.62
	One-ASQP	44.40	53.50	26.70	NA
	Paraphrase	45.70	51.00	33.00	39.60
	Ours	**50.74**	**57.29**	**36.82**	**43.50**
Rest	Extract-Classify	44.96	34.66	23.86	33.70
	SGTS-ASQE	56.02	39.59	11.08	29.61
	One-ASQP	66.30	64.20	31.10	NA
	Paraphrase	65.40	53.30	45.60	49.20
	Ours	**69.81**	**60.13**	**47.52**	**52.65**

Table 4: Test F1 scores (%) with explicit/implicit aspects and opinions on benchmark Cai et al. (2021). We compare baselines that have conducted this experiment.

age, suggesting the AGDC is necessary to construct the direct connection between explicit and implicit sentiment elements, as the syntactic dependency tree and the AMR graph only reveal the interaction between explicit sentiment element. **w/o CLN** means that we remove the conditional layer normalization but resulting in an average decline of 2.52% on F1 scores, which signifies the CLN can effectively refine the mutual indication effect between aspect-opinion pairs, as aspects provide aspect category information for opinions and conversely convey sentiment polarity information. **w/o Triaffine** indicates we simply concatenate the syntactic and semantic aware word-pair representations without using the triaffine mechanism and obtain 1.54% F1 scores degradation. Thus, it fails to align and integrate heterogeneous word-pair relations to capture higher-order interactions between sentiment elements. Overall, each module contributes to the entire performance on ASQP task.

In addition, we also conduct linguistic feature ablation to verify the role of the syntactic dependency tree and AMR graph. From the experimental results, we can conclude that these linguistic features play an indispensable role in complementing the word-pair representations from syntactic and semantic views, thereby enhancing the association between sentiment elements.

5.5 Experiment on Different Quadruplets

As mentioned in the data statistics, there exists a large percentage of reviews containing implicit aspects or opinions. Following Cai et al. (2021), we conduct the experiment on different types of sentiment quadruplets to verify the effectiveness of our model in tackling these cases. The experimental results are shown in Table 4.

Overall, we can conclude that our model significantly outperforms all baselines under different cases. Specifically, the pipeline method Extract-Classify achieves limited performance as it first obtains aspect-opinion pairs and then assigns category-sentiment labels, which overlooks the correlation between subtasks. Our model improves F1 scores by a large margin compared with tagging based methods SGTS-ASQE and One-ASQP, as their tagging scheme cannot tackle the situation where a sentence with multiple implicit expressions. Our model also exceeds the generative method Paraphrase with a significant improvement. We suppose the reason that our model based on the unified grid tagging scheme can effectively construct associations between explicit and implicit sentiment elements with syntactic and semantic linguistic features enhancement.

5.6 Case Study

A case study is illustrated in Table 5. Specifically, Extract-Classify only correctly extracts a quadruplet as it first obtains aspects and opinions and then pairs them to assign category-sentiment labels, thus leading to the error propagation problem. For MvP and MUL-ST-Scorer, they struggle to extract the quadruplets with implicit expression. We reckon these models are constrained in capturing reciprocity between sentiment elements. For One-ASQP, although achieving promising results, it cannot solve the case of the implicit aspect with implicit opinion due to the limitation of the tagging scheme. In contrast, our model based on the unified grid tagging scheme can effectively construct associations between sentiment elements with syntactic and semantic linguistic features enhancement.

5.7 Impact of the AGDC module

To intuitively reflect the impact of the adaptive graph diffusion convolution (AGDC) network, as shown in Figure 4, we further visualize word-pair representations of the syntactic dependency tree and AMR graph to compare the differences before and after the module operation.

Example	Expensive except cheese bread but nice service, I will go again.
Ground truth	(null, food#prices, expensive, negative) (cheese bread, food#prices, null, positive) (service, service#general, nice, positive) (null, restaurant#general, null, positive)
Extract-Classify	(cheese bread, food#prices, expensive, negative) ✗ (service, service#general, nice, positive) ✓ () ✗, () ✗
MvP	(null, food#prices, expensive, negative) ✓ (bread, food#quality, nice, positive) ✗ () ✗, () ✗
One-ASQP	(null, food#prices, expensive, negative) ✓ (cheese bread, food#prices, null, positive) ✓ (service, service#general, nice, positive) ✓ () ✗
MUL-ST-Scorer	(cheese bread, food#prices, null, positive)✓ (service, service#general, nice, positive)✓ () ✗, () ✗
Ours	(null, food#prices, expensive, negative) ✓ (cheese bread, food#prices, null, positive) ✓ (service, service#general, nice, positive) ✓ (null, restaurant#general, null, positive) ✓

Table 5: Case study. Marker ✓ and ✗ denotes correct and incorrect prediction.

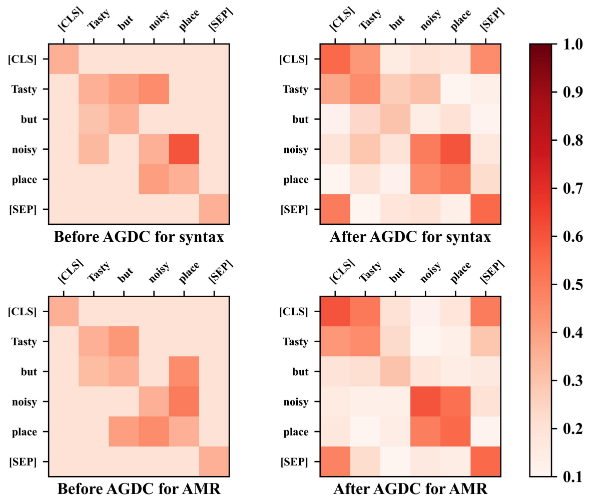

Figure 4: Visualization of syntactic dependency tree and AMR graph after the adaptive graph diffusion convolution (AGDC) operation.

Specifically, we illustrate the sampled sentence "Tasty but noisy place" containing three sentiment quadruplets, i.e., (null, food#quality, Tasty, positive), (place, ambience#general, noisy, negative) and (null, restaurant#general, null, negative). Note that we utilize the unified grid tagging scheme by adding two special tokens [CLS] and [SEP] to represent the potential implicit aspect and opinion. From the visualization we can observe that the initialized word-pair representations of syntactic and semantic structure only construct the associations between explicit sentiment elements. After the AGDC operation, it establishes the interactions between explicit and implicit sentiment elements, which is crucial for extracting sentiment quadruplets with implicit expressions.

6 Conclusion

In this paper, we propose a simple yet effective unified grid tagging scheme to formulate the ASQP task as a unified table-filling task to model word-pair relations, with two special tokens to represent potential implicit aspects and opinions. Based on this, we leverage the syntactic dependency tree and AMR graph to enrich the association between sentiment elements. To exploit these linguistic features, we first introduce the adaptive graph diffusion convolution network to construct the direct connection between explicit and implicit sentiment elements. Then we employ the conditional layer normalization to refine the mutual indication effect between aspect and opinion terms. Finally, we utilize the triaffine mechanism to align and aggregate heterogeneous word-pair relations to capture higher-order interactions between sentiment elements. Experimental results on benchmark datasets reveal the effectiveness and robustness of our model, which consistently outperforms existing methods.

Limitations

While our model based on the unified grid tagging scheme achieves promising performance, it also suffers from the following limitations:

- Due to the superiority of the unified grid tagging scheme, it can simultaneously solve subtasks such as aspect term extraction, opinion term extraction, aspect-opinion pair extraction, and aspect sentiment triplet extraction without any modifications. These subtasks are the composition of the aspect sentiment quad prediction task and contribute to the extraction of sentiment elements. Thus, exploring multi-task learning for training in conjunction with these subtasks is the potential direction for improvement.

- The implementation of our approach requires the information derived from the syntactic dependency tree and AMR graph, which affected by the parsing quality of the corresponding parsing toolkit. The good news is that the spaCy toolkit has proven its effectiveness in parsing syntactic dependency trees and AMR graphs, thus we can utilize the toolkit to support our research work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62272188) and the Fundamental Research Funds for the Central Universities 2662021JC008. We appreciate the insightful comments and constructive feedback from the anonymous reviewers. We also thank the Huazhong Agricultural University for providing the experimental resources.

References

Xiaoyi Bao, Xiaotong Jiang, Zhongqing Wang, Yue Zhang, and Guodong Zhou. 2023a. Opinion tree parsing for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 7971–7984. Association for Computational Linguistics.

Xiaoyi Bao, Zhongqing Wang, Xiaotong Jiang, Rong Xiao, and Shoushan Li. 2022. Aspect-based sentiment analysis with opinion tree generation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 4044–4050. ijcai.org.

Xiaoyi Bao, Zhongqing Wang, and Guodong Zhou. 2023b. Exploring graph pre-training for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, pages 3623–3634. Association for Computational Linguistics.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 340–350. Association for Computational Linguistics.

Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 2974–2985. Association for Computational Linguistics.

Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020. Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 6515–6524. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Timothy Dozat and Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017*.

Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. Target-oriented opinion words extraction with target-fused neural sequence labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 2509–2518. Association for Computational Linguistics.

Lei Gao, Yulong Wang, Tongcun Liu, Jingyu Wang, Lei Zhang, and Jianxin Liao. 2021. Question-driven span labeling model for aspect-opinion pair extraction. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 12875–12883. AAAI Press.

Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. 2022. LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 7002–7012. International Committee on Computational Linguistics.

Michael Wayne Goodman. 2020. Penman: An open-source library and tool for AMR graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, ACL 2020, Online, July 5-10, 2020*, pages 312–319. Association for Computational Linguistics.

Zhibin Gou, Qingyan Guo, and Yujiu Yang. 2023. Mvp: Multi-view prompting improves aspect sentiment tuple prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 4380–4397. Association for Computational Linguistics.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*.

Mengting Hu, Yinhao Bai, Yike Wu, Zhen Zhang, Liqi Zhang, Hang Gao, Shiwan Zhao, and Minlie Huang. 2023. Uncertainty-aware unlikelihood learning improves generative aspect sentiment quad prediction. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 13481–13494. Association for Computational Linguistics.

Mengting Hu, Yike Wu, Hang Gao, Yinhao Bai, and Shiwan Zhao. 2022. Improving aspect sentiment quad prediction via template-order data augmentation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 7889–7900. Association for Computational Linguistics.

Mengting Hu, Shiwan Zhao, Honglei Guo, Chao Xue, Hang Gao, Tiegang Gao, Renhong Cheng, and Zhong Su. 2021. Multi-label few-shot learning for aspect category detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 6330–6340. Association for Computational Linguistics.

Jieyong Kim, Ryang Heo, Yongsik Seo, SeongKu Kang, Jinyoung Yeo, and Dongha Lee. 2024. Self-consistent reasoning-based aspect-sentiment quad prediction with extract-then-assign strategy. *arXiv preprint arXiv:2403.00354*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7871–7880. Association for Computational Linguistics.

Kun Li, Chengbo Chen, Xiaojun Quan, Qing Ling, and Yan Song. 2020. Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7056–7066. Association for Computational Linguistics.

Songda Li, Yunqi Zhang, Yuquan Lan, Hui Zhao, and Gang Zhao. 2023. From implicit to explicit: A simple generative method for aspect-category-opinion-sentiment quadruple extraction. In *International Joint Conference on Neural Networks, IJCNN 2023, Gold Coast, Australia, June 18-23, 2023*, pages 1–8. IEEE.

Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2999–3007. IEEE Computer Society.

Ruibo Liu, Jason Wei, Chenyan Jia, and Soroush Vosoughi. 2021. Modulating language models with emotions. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 4332–4339. Association for Computational Linguistics.

Ilya Loshchilov, Frank Hutter, et al. 2017. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, 5.

Yue Mao, Yi Shen, Jingchao Yang, Xiaoying Zhu, and Longjun Cai. 2022. Seq2path: Generating sentiment tuples as paths of a tree. In *Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 2215–2225. Association for Computational Linguistics.

Mohammad Ghiasvand Mohammadkhani, Niloofar Ranjbar, and Saeedeh Momtazi. 2024. E2tp: Element to tuple prompting improves aspect sentiment tuple prediction. *arXiv preprint arXiv:2405.06454*.

Thi-Nhung Nguyen, Hoang Ngo, Kiem-Hieu Nguyen, and Tuan-Dung Cao. 2023. A self-enhancement multitask framework for unsupervised aspect category detection. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 8043–8054. Association for Computational Linguistics.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8600–8607. AAAI Press.

Joseph Peper and Lu Wang. 2022. Generative aspect-based sentiment analysis with contrastive learning and expressive structure. In *Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 6089–6095. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia V. Loukachevitch, Evgeniy V. Kotelnikov, Núria Bel, Salud María Jiménez Zafra, and Gülsen Eryigit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, pages 19–30. The Association for Computer Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 486–495. The Association for Computer Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Guixin Su, Mingmin Wu, Zhongqiang Huang, Yongcheng Zhang, Tongguan Wang, Yuxue Hu, and Ying Sha. 2024. Refine, align, and aggregate: Multi-view linguistic features enhancement for aspect sentiment triplet extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 3212–3228, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.

Siyu Tang, Heyan Chai, Ziyi Yao, Ye Ding, Cuiyun Gao, Binxing Fang, and Qing Liao. 2022. Affective knowledge enhanced multiple-graph fusion networks for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 5352–5362. Association for Computational Linguistics.

An Wang, Junfeng Jiang, Youmi Ma, Ao Liu, and Naoaki Okazaki. 2023. Generative data augmentation for aspect sentiment quad prediction. In *Proceedings of the The 12th Joint Conference on Lexical and Computational Semantics, *SEM@ACL 2023, Toronto, Canada, July 13-14, 2023*, pages 128–140. Association for Computational Linguistics.

Qianlong Wang, Zhiyuan Wen, Qin Zhao, Min Yang, and Ruifeng Xu. 2021. Progressive self-training with discriminator for aspect term extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 257–268. Association for Computational Linguistics.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020a. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585.

Zhen Wu, Fei Zhao, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2020b. Latent opinions transfer network for target-oriented opinion words extraction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9298–9305. AAAI Press.

Xiancai Xu, Jia-Dong Zhang, Rongchang Xiao, and Lei Xiong. 2023. The limits of chatgpt in extracting aspect-category-opinion-sentiment quadruples: A comparative analysis. *arXiv preprint arXiv:2310.06502*.

Yongxin Yu, Minyi Zhao, and Shuigeng Zhou. 2023. Boosting aspect sentiment quad prediction by data augmentation and self-training. In *International Joint Conference on Neural Networks, IJCNN 2023, Gold Coast, Australia, June 18-23, 2023*, pages 1–8. IEEE.

Zheng Yuan, Chuanqi Tan, Songfang Huang, and Fei Huang. 2022. Fusing heterogeneous factors with triaffine mechanism for nested named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 3174–3186. Association for Computational Linguistics.

Mao Zhang, Yongxin Zhu, Zhen Liu, Zhimin Bao, Yunfei Wu, Xing Sun, and Linli Xu. 2023. Span-level aspect-based sentiment analysis via table filling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 9273–9284. Association for Computational Linguistics.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021a. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 9209–9219. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021b. Towards generative aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 2: Short Papers), Virtual Event, August 1-6, 2021*, pages 504–510. Association for Computational Linguistics.

Wenyuan Zhang, Xinghua Zhang, Shiyao Cui, Kun Huang, Xuebin Wang, and Tingwen Liu. 2024a. Adaptive data augmentation for aspect sentiment quad prediction. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11176–11180. IEEE.

Yice Zhang, Jie Zeng, Weiming Hu, Ziyi Wang, Shiwei Chen, and Ruifeng Xu. 2024b. Self-training with pseudo-label scorer for aspect sentiment quad prediction. *arXiv preprint arXiv:2406.18078*.

Junxian Zhou, Haiqin Yang, Yuxuan He, Hao Mou, and Junbo Yang. 2023. A unified one-step solution for aspect sentiment quad prediction. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 12249–12265. Association for Computational Linguistics.

Linan Zhu, Yinwei Bao, Minhao Xu, Jianxin Li, Zhechao Zhu, and Xiangjie Kong. 2023. Aspect sentiment quadruple extraction based on the sentence-guided grid tagging scheme. *World Wide Web (WWW)*, 26(5):3303–3320.

A Appendix

A.1 Dataset Statistics

Table 7 presents the data statistics of four benchmark datasets for the ASQP task. To ensure fairness, we follow the experimental setting in Cai et al. (2021) and Zhang et al. (2021a) to divide the original datasets into the training, development and testing sets.

A.2 Implementation Details

We utilize the DeBERTa-V3-base[6] as the backbone for our model. During training, we set the number of training epochs to 100 with dropout rate of 0.5 and the batch size to 16. AdamW optimizer (Loshchilov et al., 2017) is used with a learning rate of 2e-5 for DeBERTa fine-tuning and 1e-3 for the other trainable parameters. For the AGDC module, the dimension of node feature is set to 300 and the trainable neighborhood radius t is initialized to 5.0 for each word. For the focal loss, α_t is 1.0 for "N" tag and 2.0 for other sentiment-related tags. The focusing parameter β is set to 2.0. For each dataset, we select the model with the best F1 scores on the development set and report the average results of five runs with different random seeds. Our model contains around 206M trainable parameters trained on a single NVIDIA A100 GPU with CUDA 11.0 and PyTorch 1.7.1, and the average runtime is about 0.9 seconds per batch.

[6] https://huggingface.co/microsoft/deberta-v3-base

Task	Laptop	Rest	Rest15	Rest16
ATE	83.55	87.82	84.36	85.30
OTE	83.02	86.61	83.84	86.76
AOPE	70.96	77.18	71.41	75.19
ASTE	64.10	71.54	67.05	71.39

Table 6: Test F1 scores (%) on several subtasks.

A.3 Experiments on Subtasks

To further demonstrate the effectiveness of our model based on our proposed unified grid tagging scheme, we conduct experiments on four subtasks of the ASQP, i.e., aspect term extraction (ATE), opinion term extraction (OTE), aspect-opinion pair extraction (AOPE) and aspect sentiment triplet extraction (ASTE). The experimental results are shown in Table 6. Note that our method can simultaneously tackle these subtasks without additional modifications when addressing the ASQP task.

Specifically, we observe that our model achieves promising results on these subtasks. We suppose the reason that our unified grid tagging scheme can effectively construct the associations between sentiment elements by modeling word-pair relations. Furthermore, we can find that the extraction performance decreases on a more complex subtask with the increase of sentiment elements as it causes the increasing probability of mismatching and missing sentiment elements.

A.4 Potential Practical Applications

The time complexity of our model is quadratic relative to the input data. The primary source temporal overhead in our model stems from two aspects, one is the transformer's attention operations and the other is the inefficiency of the matrix exponential calculation using in adaptive graph diffusion convolution module. For the second aspect, we can utilize the `torch.linalg.matrix_exp()` function to speed up the operation by upgrading PyTorch to version 1.9 or above.

As for space complexity, our model takes up an additional parameter space occupation owing to the

Dataset		#S	#Q	#C	#POS	#NEU	#NEG	EA&EO	EA&IO	IA&EO	IA&IO
Laptop	**Train**	2934	4172	121	2583	227	1362	2343	895	681	253
	Dev	326	440		279	24	137	259	93	65	23
	Test	816	1161		716	65	380	673	253	169	66
Rest	**Train**	1530	2484	13	1656	95	733	1662	215	374	233
	Dev	171	261		180	12	69	173	28	34	26
	Test	583	916		667	44	205	596	107	122	91
Rest15	**Train**	834	1354	13	1005	34	315	1082	-	272	-
	Dev	209	347		252	14	81	287	-	60	-
	Test	537	795		453	37	305	577	-	218	-
Rest16	**Train**	1264	1989	13	1369	62	558	1543	-	446	-
	Dev	316	507		341	23	143	403	-	104	-
	Test	544	799		583	40	176	620	-	179	-

Table 7: Statistics for different datasets. #S, #Q and #C mean the total number of sentences, quadruplets and aspect categories. #POS, #NEU and #NEG denote the number of positive, neutral and negative sentiment quadruplets respectively. EA, EO, IA and IO denote explicit aspect, explicit opinion, implicit aspect and implicit opinion.

word-pair relations construction, which is notably minor when compared to the parameter size of the pre-trained model.

Based on the above description, our model demonstrates the commendable scalability for practical applications.